Automatic Analysis and Musicological Interpretation of Human Free Sorting of Musical Excerpts

Nicolas Dauban,
Christine Sénac,
Julien Pinquier

IRIT
University of Toulouse, UPS
Toulouse, France
nicolas.dauban@irit.fr
christine.senac@irit.fr
julien.pinquier@irit.fr

Pascal Gaillard,
Ludovic Florin

CLLE
University of Toulouse, UT2J
Toulouse, France
pascal.gaillard@univ-tlse2.fr
ludovic.florin@univ-tlse2.fr

Paul Albenge

IREMUS
University of Paris-Sorbonne
Paris, France
paul.albenge@gmail.com

***Abstract*—Most content-based music recommendation systems are relying on audio features which do not always match with musicological criteria. This paper describes the experimental protocol and the results of a sorting experiment, which leads to an 'average categorization' by volunteers. An automatic analysis afterward aims at identifying relevant acoustic parameters based on the obtained categories and sub-categories. A musicological analysis was also done in parallel.**

Keywords–Categorization; Music Information Retrieval; Recommendation.

I. Introduction

The role of music recommendation algorithms is to offer new songs to users of online music listening platforms. Research in Music Recommendation (MR) is very recent and underdeveloped in the academic world, because of the limitation -due to licensing issues- of access to the music signal on a large scale. Because basing a recommendation on simple metadata from collaborative filters is not always relevant, more and more works are based on the expertise acquired in Music Information Retrieval (MIR), which aims to extract information from the signal at different scales (notes, chords, sequence of notes, etc.) in order to characterize for example an instrument or to calculate descriptors, such as the tempo or the main melody. See [1] for a state of art on the MIR.

Thus, some authors have tried to rely on the measure of similarity between pieces of music [2]. While this approach is relevant for genre classification (the closest task to musical recommendations), it is quite disappointing for MR. Also, it was natural to introduce, in parallel to content-based approaches, information about user preferences [3] [4], or user behavior [5]. However, whatever the method, less known pieces (located in the 'long tail') are never (or rarely) proposed: some works aim to remedy this problem [6] [7].

One of the main difficulties raised by content-based methods is the selection of parameters. Indeed, among all the parameters we can extract from an audio signal, which of them can describe and explain the listeners' liking? How can we link these acoustic parameters to musicological and perceptive criteria? These problems are the major issues of the project in which this free categorization experience fits. The purpose of this experiment is to identify both acoustic parameters and musicological or 'non-expert' criteria according to which the subjects classify the pieces, starting from the assumption that the tastes of a listener are linked to a fixed combination of parameters or criteria.

Section II describes the constitution of the corpus and the experimental conditions. Section III presents the data generated by the experiment and the way we processed it automatically and how it was interpreted with a musicological point of view. Section IV describes a method based on audio features which aims at reconstruct the categorization made by volunteers.

II. Experimental protocol

A. Corpus with musicological criteria

One of the first steps of the project was to build a corpus, which had to meet several requirements: (1) wide range of musical genres; (2) good quality excerpts: Audio CD (stereo, 16 bits, 44.1 kHz); (3) long enought excerpts (at least 20s) and in sufficient numbers; (4) preferably with a copyright-free access database.

The corpus has been built with a musicological approach. First, we made a set of 15 criteria which can define the music in the most comprehensive way possible without using a commercial classification like genres. The concepts and lexicon used here rely mainly on the work of Pierre Boulez and Gilles Deleuze in [8] and [9].

- Recording Quality: perception of the support and mean of recording (noises, sound spectrum, intensity, etc.).
- Prevalence of an Instrument: salience of a special timbre.
- Voice: presence or absence of voice, type of voice (spoken, sung, declarative, repetitive, etc.).
- Space: feeling and representation of a diffusion space, deepness of the musical field.
- Memory Work: presence of one or several memorable elements, repetition of an element (stricte or similar), clear perception of a pattern or logic.
- Dynamic: change of quantity/density of events, contrast in the musical development.
- Narrative Development: evolution of musical elements, presence of different parts relatively distinct.

- Smooth/Striated time: according to Boulez [10], presence or absence of beat, diversity of elements, variation in quantity of elements in a short moment.
- Sensorimotor: instrumentalists' music, mostly animated by a desire of gesture and a research of the sensorial effect of the sound. As exemple, African music, percussion improvisations, jazz solos, or concrete pieces of music by Pierre Henry.
- Representation: what represents by a visible or hidden way the reality on the plastic plan, by trajectories, speeds, impacts or event realistic noises (Gregorien, occidental romanticism, many contemporary music).
- Rules: any written music, whether written as a classical counterpoint or transmitted orally as Pygmy polyphonies or M'Baka horns.
- Energy: intensity, body implication, involvement of the musician.
- Level of Technicity: there are two levels, instrumental and composition. Perception of the assurance of musician's intentions and/or presence of structural concepts.
- Cultural Elements: reference to a socio-cultural class.
- Chronological Situation: perception of elements specific to an era, such as type of recording, type of play, reference to a particular aesthetic.

For each of these criteria, we empirically selected three significant pieces which contain different musical characteristics in order to propose an eclectic ensemble. Although these would have been selected to initially correspond to a specific criterion, it is possible to find characteristics of other criteria. The goal here is not to recover this classification in the experimental results but to see which criteria were particularly relevant in the volunteer's sorting. For each selected track, we had to select a short excerpt in order to keep the experiment from being too long for the volunteers. Excerpts had to be still relevant regarding to the corresponding criterion. In the end, a corpus of 45 excerpts of 20 seconds was defined (see Figure 1).

B. Experimental conditions

In order to limit the impact of the age of the participants on the results, we used participants/volunteers from 20 to 25 years old (30 in number). For the experiment, we used the TCL-labX tool [11] [12].

The interface (see Figure 2) was presented in an identical way to all volunteers who were asked to freely sort excerpts and thus form as many categories as they wished, based on the similarities between the pieces. To do so, the users could listen as many times as necessary excerpts and could move and group them freely on the interface.

III. Free sorting and interpretation

A. MetaData

For each volunteer, the program generates a file in which is indicated the distribution of the excerpts in the different classes. The software also generates a "cookie" file containing the history of the operations performed by the user: moving icons and listening to excerpts. We can replay all actions performed by the volunteer. In addition, the software carries out an automatic analysis of these files in order to extract several statistics on the participants. The average duration

Excerpt	Begining	Artist - Title
1	00:00	Les Doubles Six - Au Bout du Fil (Meet Benny Bailey)
2	00:02	Bill Bruford - One Of A Kind (Part 1)
3	00:00	Harry Burleigh - Go Down Moses
4	03:20	Rautavaara - Cantus Arcticus 2e mvt
5	00:50	Hector Berlioz - Symphonie Fantastique, Op. 14, Songe d'une nuit de sabbat
6	00:00	Corette - Concerto pour musette de cour 2 Adagio
7	00:00	Namibie Chant De Guerison - Nom Tzisi
8	00:00	Han Bennink & Willem Breuker - Mr. M.A. de R. in A.
9	00:00	Death Grips - Thru The Walls
10	00:00	Jazzoo - le pic et le moineau
11	00:24	Big Satan - Geeza
12	00:02	Lords Of The Underground - Here Come The Lords
13	00:02	Pygmées Aka
14	00:00	Deux Chants De Jeu Et De Danse - Polynésie Occ.
15	00:10	James Brown - Mother Popcorn
16	03:00	Suisse Yodel - Zauerli
17	00:13	Horace Silver - Capverdian Blues
18	03:10	Awa Poulo - Dimo Yaou Tata
19	00:00	Pharoah Sander - Love Will Find a Way
20	00:00	David Fiuczynski - Moonring Bacchanal
21	00:00	André Minvielle - L'Alambic
22	00:25	Edgard Varèse - Un Grand Sommeil Noir
23	00:30	The Residents - This Is Man's World
24	00:28	Theo Bleckmann & Ben Monder Late Green
25	03:40	Aphex Twin - Circlont14 [Shrymoming Mix]
26	00:00	Naked City - Une Correspondance
27	05:00	Bugge Wesseltoft - Dreaming
28	00:40	Ali Farka Touré - Sabu Yerkoy
29	00:46	A Ram Sam Sam
30	00:05	Sleepytime Gorilla Musuem - The Putrid Refrain
31	08:20	John Zorn - Through The Night
32	00:20	Liadov : Baba Yaga
33	00:28	Don Ellis - Strawberry Soup
34	00:30	Ligeti - Quatuor à cordes n°2 - come un meccanismo di precisione
35	00:27	Arvo Pärt "Ludus" du Tabula Rasa
36	00:20	John Zorn, Filmworks - Cynical Hysterie Hour Through the
37	01:50	Jaco Pastorius - Come On, Come Over
38	00:00	Bach/ Glenn Gould - The Art of the Fugue, BWV 1080- Contrapunctus III
39	02:30	Julien Loureau - Conrod
40	00:25	Dayton - Krackity Krack
41	00:43	Aka Moon - For Drummers Only
42	00:15	Sec - Run Away
43	00:23	Tool - Lateralus
44	00:00	Bruckner - scherzo 9e symphonie
45	00:30	Naked City - Speedball

Figure 1. List of 20 seconds excerpts and their beginning time.

Figure 2. Interface presented at the beginning of the experiment (each excerpt is represented by a numbered icon), and after the free sorting (excerpts are grouped by volunteer).

of the experiment was 37 minutes, the maximum duration exceeded one hour (1h 2min) and the minimum duration was 15 minutes. The standard deviation over the duration of the experiment is 10 minutes. On average, participants formed 15.5 classes, the minimum being 8 classes and the maximum 20. The standard deviation on the number of classes is 3.2.

B. Automatic Results Analysis

1) Matrices of co-occurrence and dissimilarity: to accomplish this task, we based ourselves on the work of [13]. First, we built a co-occurrence matrix C^i for each participant i. A co-occurence matrix is square and symmetrical, of dimension $N \times N$ with N equal to the number of sorted excerpts. In each cell, we indicate the distance between two excerpts: if these two excerpts are in the same category, the distance is considered as zero, otherwise we assign a unit distance.

Then, we calculate an average co-occurrence matrix of all participants, called the D dissimilarity matrix. This dissimilarity matrix gives us a distance measurement for each pair of

musical excerpts, this distance being based on the classification by the n participants.

We have also computed a matrix of variance M_{var} from the matrices of co-occurrence C^i and the dissimilarity matrix D (see equation 1). For a pair of excerpts, this variance is null if the set of n participants sorted these two excerpts identically. Conversely, this variance is maximal ($var_{max} = 1$) if half of the participants put these excerpts together, and the other half separately.

$$M_{var}^{j,k} = \frac{1}{n} \sum_{i=1}^{n} (C^{i,j,k} - D^{j,k})^2 \quad (1)$$

with $M_{var}^{j,k}$ the cell of the line j and column k of the matrix M_{var}, $C^{i,j,k}$ the cell (j,k) of the matrix C^i and $D^{j,k}$ the cell (j,k) of the matrix D.

Since the C^i matrices contain only binary values, for a pair of excerpts, a null variance necessarily corresponds to a dissimilarity of 1 or 0, and a unit variance necessarily corresponds to a dissimilarity of 0.5. The closer the dissimilarity is to an extreme value (1 or 0), the more these two excerpts will be 'unanimously' put in their class by volunteers. The closer the dissimilarity is to 0.5, the more these two extracts will have 'divided the opinion' of the volunteers.

2) Dendrogram: from the dissimilarity matrix, we were able to perform an ascending hierarchical classification. We have chosen to use the *Ward's method* [14], which consists in grouping the classes so that the increase of the inertia interclasses, given by (2) is maximized. This, according to the *Huygens theorem*, is equivalent to minimize the increase of the intraclass inertia (see (3)) [15].

$$I_e = \frac{1}{n} \sum_{i=1}^{k} n_i \times d^2(g_i, g) \quad (2)$$

$$I_a = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} d^2(e_j, g_i) \quad (3)$$

We obtained the dendrogram of Figure 3. It tells us about the links between excerpts put by users. The vertical axis represents the distance between excerpts or groups of excerpts. Thus, two excerpts that have been very often placed together by the volunteers have a low link on the figure, such as on the numbers 11 and 17 or the numbers 15 and 37.

C. Musicological Interpretation

The dendrogram obtained previously was interpreted from a musicological point of view in order to understand how the volunteers had made their classification. The dendrogram has been annotated with the criteria common to excerpts belonging to the same branch, see Figure 3.

The name indicated under each node corresponds to a presumed musicological criterion common to all the excerpts which are below it. This name was choosen analysing the content of each category. The four main categories are described as follows.

1) Audio-Tactile: Pieces of music in this category are all very rhythmic and belong to the genre jazz or funk and more generally to Africo-American. "Audio-tactility" refers to a particular relationship with the body. Within this category, the excerpts have been distinguished by the predominant instrument.

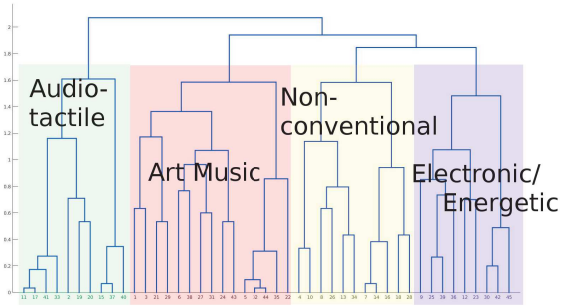

Figure 3. Dendrogram annotated by musicologists. y axis corresponds to the distance between two excerpts or clusters of excerpts; excerpts to the x axis.

2) Art Music: this category is a hybrid grouping of academic Western Music, of music evoking a sacred or spiritual aspect and finally of the excerpts where the voice is predominant.

3) Non-conventional: this category includes music built outside the conventional rules governing Western music, particularly based on melody and harmony. Therefore, we find music without pitch of precise notes and the non-Western music. Any music that does not follow the hierarchies present in Western codes is necessarily perceived as a group apart.

4) Electronic/Energetic: most of the excerpts in this category were produced from electronic instruments. This category also includes pieces with contrasts in terms of energy. However, the distance remains particularly high between these two sub-categories.

The objective was not to recover, via the results of the experiment, the free classification created but to see which dimensions are the most important while listening to music. We can notice that the musicological criteria used to establish the corpus are not found through the free classification of non-expert participants. This shows that there are different types of musical analysis and valuation. The musicological criteria allowed us to obtain a very varied corpus and the categories identified by the participants reveal other more accessible criteria for non-experts.

IV. Towards an automatic classification

The aim here was to verify whether an automatic classification of musical extracts based on acoustic parameters could approach the free classification made by humans and based not only on the signal, but also on knowledge. As music (in term of production) commonly refers to a series of sound events (notes, percussive sounds, voiced or unvoiced sounds) defined by their rythm, timbre, dynamics, and pitch, we have extracted some of these parameters. So, we calculated 31 audio parameters on each excerpt of the corpus using the MIR Toolbox [16]. The extraction of these parameters is described in detail in the following subsection.

A. Extraction of Acoustic Parameters

1) Rhythm: it describes the temporal location of sound events and their duration. Generaly, in conventional western music, a regular pulse determines the beat, a measure being composed of several beats. In a score, the rhythm (inside the beat) is described by the different shapes of notes (crotchet, quaver, semiquaver, etc.) and of silences (pause, minim, etc.) as well as by the time signature.

- Event Detection and Density: all the rhythmic parameters are based initially on the temporal location of each event. For this, we use a peak detection algorithm on the signal envelope. Once the peaks are detected, we can then calculate the number of events per second. These features are extracted on 10 seconds window without overlapping.
- Tempo: the tempo calculation, which is based on a detection of the periodicity of events, selects the highest peak. Periodicity detection is performed using the autocorrelation function [16]. This feature is extracted on a 3 second window with and overlap of 0.3 seconds.
- Pulse Clarity: the pulse clarity can be calculated according to the method detailed in [17]. This parameter describes how much the beat is dominant in the rhythm, or in other words, how much emphasis is placed on the beats: for example, the clarity of the beat is strong for disco rhythms, and is often low for complex rhythms, like those of jazz. This feature is extracted on a 5 second window with and overlap of 0.5 seconds.

2) Timbre: it describes the spectral composition of a note, that is to say the amplitude of the harmonics and the variation in time of these harmonics. This distinguishes, for example, two notes played at the same pitch by a piano and a guitar. These features (excepted the attack) are extracted on a 50ms window with and overlap of 25ms.

- Attack: the attack of a note describes the variation of amplitude at the moment when it is played. It is measured by its duration, its amplitude or by its slope [18]. For example, struck strings of the piano have a stronger attack than violin strings played with a bow. When a note is detected, the beginning and the end of the attack are marked, then the difference of amplitude or the duration between the two points is calculated (the slope is obtained using these two informations).
- Zero Crossing Rate: it is calculated on the original signal by multiplying all successive pairs of samples, and iterating a variable when the product is negative (signal change). This variable is then divided by the duration to obtain the rate [19].
- Rolloff Frequency: it informs us about the amount of energy present in the low frequencies. On a spectrum, we calculate the frequency below which 85% of the energy is contained. The lower the frequency, the more energy is concentrated in the low frequencies.
- Brightness: it informs us about the amount of energy present in the high frequencies [20]. On a spectrum, we calculate the amount of energy present beyond a fixed frequency (usually 1500 Hz).
- Statistical Parameters of Spectral Distribution: it is possible to calculate statistics as well as moments of different orders on the spectrum, such as Centroid, Spread, Skewness, Kurtosis, Flatness, as well as the Entropy.
- MFCC (Mel Frequency Cepstral Coefficients): MFCCs are cepstral coefficients calculated by a discrete cosine transform applied to the power spectrum of a signal [21]. The different frequency bands are determined according to the perceptive logarithmic scale Mel, which is modeled on the human hearing system.
- Roughness: it describes the phenomenon of audible beat in the presence of two near frequencies [22]. Two notes spaced half tone apart (or less) will generate strong roughness, which decreases as spacing increases. Roughness is almost zero from 5 half tones.
- Irregularity of a Spectrum: this is the degree of amplitude variation of two successive peaks (harmonics or not) of the spectrum [16].

3) Dynamics: it describes the relative amplitude of different sounds, which results in shades of intensity. On a score, the dynamics is indicated by terms, such as 'pianissimo' or 'forte' that tell the musician to play relatively more or less loudly. In signal processing and generally, dynamics describes the range of variation of the different values taken by a signal. In music, dynamics describes the ratio of sounds of strong and weak amplitudes. These features are extracted on a 50ms window with and overlap of 25ms.

- RMS (Root Mean Square) level: the effective value of an ergodic random signal over a time interval is the square root of the square signal mean, or the square root of its mean power. In practice, for a discrete time signal, the RMS level is calculated on a finite number of samples.
- Low Energy Rate: it is the number of points whose value is less than the RMS value of the signal. For a signal with peaks at the high RMS level, this rate will be high whereas for a signal at the RMS level rather constant, this rate will be low.

4) Pitch: it describes the fundamental frequency of a sound played by an instrument, which defines the note.

- Note detection: the default method for detecting notes is to decompose the signal into several frequency bands, then calculate the auto-correlation and finally detect the peaks in order to obtain an estimate of the notes. This feature is extracted on a 46.4ms window with and overlap of 10ms.
- Harmonies Detection: from the detection of notes, it is then possible to detect harmonies, that is to say combinations of different notes. It is also possible to calculate the key of an extract, as well as the temporal evolution of all these parameters. This feature is extracted on a 743ms window with and overlap of 74ms.

B. Selection of Acoustic Parameters

We averaged each parameter to obtain a matrix of the form $N \times P$ with $N = 45$ (excerpts) and $P = 31$ (parameters). Note that by keeping only the mean, we lose the temporal evolution, but this allows us to have only one scalar per excerpt and per parameter. For each parameter, we computed the distance for each pair of excerpts and thus form a dissimilarity matrix P^i for each parameter i. Then, we established a model of the matrix of dissimilarity of the human free sorting from a linear combination of the matrices of the parameters.

$$M_{model} = \sum_{i=1}^{31} a_i P^i \tag{4}$$

The values contained in each dissimilarity matrix were normalized between 0 and 1 in order to remain consistent with the matrix values of the free sorting.

Rather than using all P^i matrices, we selected the most relevant matrices by computing the correlation coefficient of each matrix of parameters with the matrix of the volunteers and we selected the m more correlated. Indeed, the matrices the most correlated with the matrix of dissimilarity established during the free sorting are by definition the most 'similar'.

C. Regression

1) Complete matrix: with these first m matrices, we used a gradient descent algorithm to find the best linear combination, the criterion to be optimized being the quadratic error between this linear combination and the dissimilarity matrix of the free sorting. This algorithm therefore returns the coefficients a_i by which the matrices of dissimilarity are multiplied in order to obtain the matrix most resembling the dissimilarity matrix formed by the set of results of the volunteers. These coefficients inform us of the importance of each parameter: if a coefficient is low then it is not influential for the volunteers to sort the pieces, and vice versa.

To simplify the calculations, the dissimilarity matrices have been transformed into V_p and V_d vectors of length $L = 45 \times 45 = 2025$.

For m used dissimilarity matrices of parameters, the quadratic error equation is defined by:

$$J = \sum_{j=1}^{L} \left[\left(\sum_{i=1}^{m} a_i V_p^{i,j} \right) - V_d^j \right]^2 \tag{5}$$

The gradient of this error is:

$$\overrightarrow{\text{grad}}\ J = \begin{bmatrix} \frac{\partial J}{\partial a_1} \\ \dots \\ \frac{\partial J}{\partial a_k} \\ \dots \\ \frac{\partial J}{\partial a_M} \end{bmatrix} \tag{6}$$

where:

$$\begin{aligned} \frac{\partial J}{\partial a_k} &= \sum_{j=1}^{L} \frac{\partial \left[\left(\sum_{i=1}^{m} a_i V_p^{i,j} \right) - V_d^j \right]^2}{\partial a_k} \\ &= 2 \sum_{j=1}^{L} \left[\left[\left(\sum_{i=1}^{m} a_i V_p^{i,j} \right) - V_d^j \right] V_p^{k,j} \right] \end{aligned} \tag{7}$$

We successively tested the algorithm with the m 'best' parameters, in the sense of the correlation. By successively increasing the number of parameters, the total squared error decreases to 2.20. From 7 parameters, the error increases again. The 6 first parameters are: *Irregularity, Brightness, Rolloff, Harmony Change Detection, Spectrum Entropy, Attack.*

From the matrix estimated with 6 parameters, we generated a new dendrogram (see Figure 4) in order to visually compare the result of this estimation with the dendrogram obtained at the end of the free sorting (see Figure 3). We can see these two dendograms do not resemble a lot one another. This can be explained by the fact that the participants did not use the same 'rule' to classify all the excerpts. For example, some excerpts have been grouped with respect to rhythmic similarities, and others with respect to their melody. It is difficult to establish a general rule on the parameters to estimate with good precision the overall classification. This is probably due to the fact that the excerpts were highly variable. If we consider dendrogram sub-parts, the excerpts belonging to each of them are more similar to each other, and we can therefore suppose that it will be easier to isolate discriminant parameters.

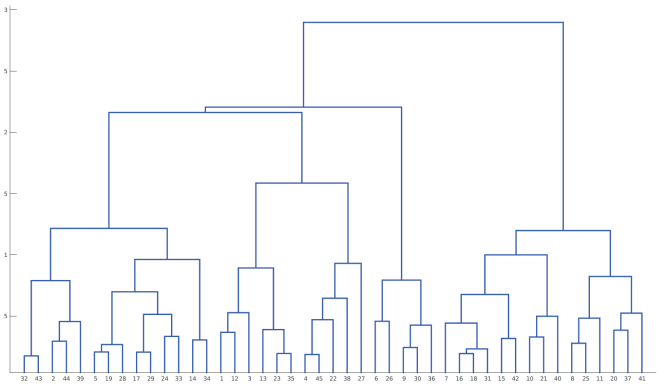

Figure 4. Dendrogram estimated from 6 parameters.

2) Sub-Matrices/Parts: we used the same method as above on each of the four sub-parts, named by the musicologists: Audio-Tactile, Art Music, Non-conventional and Electronic/Energetic. We calculated the most correlated criteria, and used them again to form linear combinations.

3) Audio-Tactile: the most correlated parameters are: *Brightness, Irregularity, MFCC10, Attack, MFCC4, Spectrum Entropy, Pulsation Clarity, Zero Crossing Rate, Low Energy, Spectrum Kurtosis, MFCC3, Rolloff, Tempo, MFCC8.*

At the end of the gradient descent, the total squared error is 5.5. We note that several MFCCs were involved in the classification. We can explain this by the fact that for this category, the volunteers distinguished the excerpts according to the predominant instruments. The dendrogram was well reconstructed, except for excerpts 5 and 6 which were exchanged.

4) Art Music: For this category, the results were rather mitigated, even using all the parameters. Indeed, at the end of the gradient descent, the total squared error is 9.5. These weaker results can be explained by the fact that this category contains more excerpts, which are ill-matched making it more difficult to generalize a categorization rule.

5) Non-conventional: We obtained good results with 18 parameters. At the end of the gradient descent, the total squared error is 3.8. The most correlated parameters are: *Sensory Dissonance, Attack, Number of Events Per Second, Zero Pass Rate, Spectrum Kurtosis, Key Clarity, Entropy, MFCC8.*

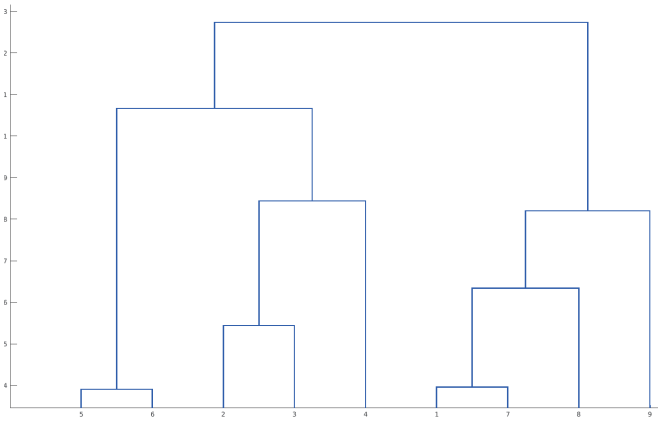

Figure 5. Dendrogram obtained for the Electronic/Energetic category.

6) Electronic/Energetic: The method has been the most efficient for this group (see Figure 5). Except for the first excerpt, the dendrogram was well reconstituted. For each sub-group i of size N_i, the initial indices of the excerpts were replaced by indices ranging from 1 to N_i. If the dendrogram of a sub-group has been reconstituted, the excerpts are placed in ascending order. We obtained a total squared error of 2.8. We used the following 9 parameters: *Attack, MFCC3, MFCC8, MFCC11, Rolloff, Brightness, MFCC4, Zero Crossing Rate, MFCC0 (i.e. Energy)*.

Overall, the results are quite satisfactory because we were able to reconstruct the dendrogram of each category with a limited number of errors.

D. Classification of new excerpts

The objective of this part was to find a method to assign to a 'new' excerpt the right category. In all the methods that follow, we successively considered each excerpt as a new individual, taking care to remove it from the learning base (leave-one-out cross-validation). The score for each method is therefore between 0 and 45 (where all the excerpts were assigned to the right categories). In addition, the parameters were centered and reduced in order to eliminate the influence of the unit of measure used for each of them.

The first method consisted in calculating the center of gravity of each category according to the 31 parameters, and then assigning the new extract to the class with its nearest barycenter: we thus obtained 30 correct assignments (67%).

In the second method, we retained a small number of parameters: we observed which parameters were relevant for the classification in the sub-categories (section IV-C2) and we kept only those which were the most correlated in the 4 classifications. They are: *Change Harmony Detection, Attack, Spectrum Entropy, Rolloff, MFCC0, and Irregularity*. We thus obtained 22 correct assignments (33%): the selected parameters were therefore not particularly relevant.

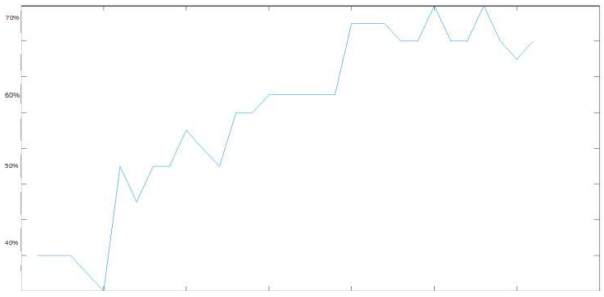

Figure 6. Attribution score based on the number of parameters used. y axis corresponds to the score obtained in function of the number of parameters (x axis).

In the third method, we used the N most correlated parameters in ranking the same way as in the section IV-C1. In Figure 6, we see that the score increases globally with the number of parameters but sometimes decreases when we use a new one. The maximum score (71%) is reached for 25 parameters : all but *3rd, 4th and 6th MFCC, the Spectral Flatness, Inharmonicity and Spectrum Kurtosis*. This method proved to be the best.

V. Conclusion and prospects

For this experiment, a corpus was built according to a wide range of musicological criteria. Different audio parameters were also computed on the excerpts of the corpus.

Volunteers have performed a free sorting task on this corpus. Analysis of the experimental results let us establish an average human classification of excerpts by volunteers, which has been represented in the form of a dendrogram in which appear four main groups with sub-groups. We noticed that these sub-groups were built according to some of the musicological criteria but also according to 'non expert' criteria such as genres.

In order to automatically reconstruct this human classification, we have established a hierarchy in the parameters relevance depending on their correlation with the volunteers' classification. We saw that this automatic reconstruction is more efficient to distinguish sub-groups within a group instead of groups between them. Finally, the identified parameters can be selected for an application in music recommendation.

References

[1] M. Schedl, E. Gómez, and J. Urbano, "Music information retrieval: Recent developments and applications," Foundations and Trends® in Information Retrieval, vol. 8, no. 2-3, 2014, pp. 127–261.

[2] B. McFee, L. Barrington, and G. Lanckriet, "Learning content similarity for music recommendation," IEEE transactions on audio, speech, and language processing, vol. 20, no. 8, 2012, pp. 2207–2218.

[3] A. Van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in Advances in neural information processing systems, 2013, pp. 2643–2651.

[4] E. J. Humphrey, J. P. Bello, and Y. LeCun, "Moving beyond feature design: Deep architectures and automatic feature learning in music informatics." in ISMIR. Citeseer, 2012, pp. 403–408.

[5] M. Schedl and D. Hauger, "Tailoring music recommendations to users by considering diversity, mainstreaminess, and novelty," in Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015, pp. 947–950.

[6] Ò. Celma Herrada, "Music recommendation and discovery in the long tail," Ph.D. dissertation, 2009.

[7] M. A. Domingues, F. Gouyon, A. M. Jorge, J. P. Leal, J. Vinagre, L. Lemos, and M. Sordo, "Combining usage and content in an online recommendation system for music in the long tail," International Journal of Multimedia Information Retrieval, vol. 2, no. 1, 2013, pp. 3–13.

[8] R. Bogue, Deleuze's way: Essays in transverse ethics and aesthetics. Routledge, 2016.

[9] G. Deleuze and F. Guattari, A thousand plateaus: Capitalism and schizophrenia. Continuum, 2004.

[10] P. Boulez, Boulez on music today (trans. by Bradshaw, Susan and Rodney Bennett, Richard). London: Faber and Faber, 1971.

[11] P. Gaillard, "Laissez-nous trier ! tcl-labx et les tâches de catégorisation libre de sons." Le sentir et le dire : Concepts et méthodes en psychologie et linguistique cognitive, 2009, pp. 189–210.

[12] http://petra.univ-tlse2.fr/tcl-labx, [Online; accessed 15-January-2019].

[13] J. P. Barthélémy and A. Guénoche, Trees and Proximity Representations. John Wiley and Sons, 1991.

[14] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," Journal of the American statistical association, vol. 58, no. 301, 1963, pp. 236–244.

[15] G. Saporta, Probabilités, analyse des données et statistique. Editions Technip, 2006.

[16] O. Lartillot, "Mirtoolbox 1.6.1 users manual," 2014.

[17] O. Lartillot, T. Eerola, P. Toiviainen, and J. Fornari, "Multi-feature modeling of pulse clarity: Design, validation and optimization." in ISMIR, 2008, pp. 521–526.

[18] J. M. Grey, An Exploration of Musical Timbre Using Computer-based Techniques. Department of Psychology, Stanford University., 1975.

[19] F. Gouyon, F. Pachet, and O. Delerue, "On the use of zero-crossing rate for an application of classification of percussive sounds," in Proceedings of the COST G-6 conference on Digital Audio Effects (DAFX-00), Verona, Italy, 2000.

[20] P. Laukka, P. Juslin, and R. Bresin, “A dimensional approach to vocal expression of emotion,” Cognition & Emotion, vol. 19, no. 5, 2005, pp. 633–653.

[21] B. Logan, “Mel frequency cepstral coefficients for music modeling.” in ISMIR, vol. 270, 2000, pp. 1–11.

[22] W. A. Sethares, Tuning, timbre, spectrum, scale. Springer Science & Business Media, 2005.